Investigating Interactions with Teleoperated and Autonomous Humanoids Using a Suit-Based VR Teleoperation Interface

Maxwell Bennett
Kindred
San Francisco, California
max@kindred.ai

Daria Thames
HRILab
Tufts University
Medford, MA
daria.thames@tufts.edu

Tom Williams
MIRRORLab
Colorado School of Mines
Golden, CO
twilliams@mines.edu

Matthias Scheutz
HRILab
Tufts University
Medford, MA
matthias.scheutz@tufts.edu

ABSTRACT

In this paper, we discuss how the use of a suit-based VR teleoperation interface and paired humanoid robot allowed us to investigate differences in how humans instruct and perceive autonomous and teleoperated humanoid robots, as well as other humans. In particular, as the use of teleoperated humanoids increases, we are interested in how physical embodiment and perceived autonomy affect how humans instruct and perceive these humanoids. We abridge our previous work and empirical results, and describe lessons learned with respect to the design and use of VR suit-based teleoperation interfaces such as the one used.

1 INTRODUCTION

It is becoming increasingly important to understand how humans will perceive and instruct robots. Knowing how humans will instruct robots is especially important both for robot developers seeking to enable learning-from-demonstration capabilities as well as those seeking to work with robots in more collaborative tasks requiring sophisticated motion or dexterity.

What is more, with the increased use of teleoperated humanoid robots, it is important to recognize whether any such differences are due to the physical embodiment or to the perceived level of autonomy of the instructee. While some previous work [9, 10] has begun to look at linguistic differences in human-to-human and human-to-robot instructions, it has not considered such possible effects. Furthermore, in that work a specific instruction task was provided by the experimenters, which may have biased the utterances used in task instructions. Little empirical research in HRI has investigated perceptions of teleoperated humanoid robots, and to the best of our knowledge all previous research investigating human perceptions of autonomous versus teleoperated robots has been *observational*, in addition to using robots controlled through graphical interfaces.

In this paper we abridge much of our previous work [1] and take a deeper focus to how using a VR teleoperation interface, in which teleoperator motions are replicated by a robot in real time, allowed us to better investigate *interaction* differences between human-to-human and human-to-robot task instructions.

2 RELATED WORK

While there has been significant research into robot teleoperation in general, little empirical research has made use of teleoperated *humanoid* robots. This is deeply problematic because humanoid robots are uniquely suited to perform many tasks in environments designed for human beings [4], and because research has shown that humanoid robots may be perceived using the same cognitive processes normally reserved for perception of human agents [7]. Furthermore, to the best of our knowledge, when the work presented in this paper was performed, all previous *empirical research* involving teleoperated robots had relied on joystick based or graphical interface based Wizard-of-Oz interfaces [4, 12].

As the complexity of humanoid robots has increased, interfaces for controlling such robots through teleoperation have moved from rudimentary 2D graphical interfaces to more complex interfaces in which robot control is based on natural human gestures and motions [3] through sensor-equipped bodysuits or video-based motion-capture devices [8, 11, 14]. Rodriguez et al., for example, present a vision-based teleoperation interface that uses motion capture data of a human operator to select and send motion commands such as walking and leaning to a NAO robot [8]. Using vision-based interfaces, however, presents a new challenge, as the difficulty of extracting accurate representation of human motions from video can be challenging in poorly lit or cluttered environments.

For teleoperated humanoid robots, it is crucial for interaction studies to use teleoperation interfaces that replicate the natural motions of human teleoperators [8, 11], as this may greatly increase the perceived human-likeness of the robot and significantly reduces the likelihood that a robot will be physically unable to comply with a given instruction.

More recently, robot developers have taken this type of teleoperation interface one step further, using virtual reality devices to provide visual immersion and head control. This not only allows

the teleoperator to visually explore their environment the way they would if they were present in that environment, but also provides remote teammates information regarding the location of the teleoperator's gaze, which is a valuable source of information both for completion of task-based goals and for engaging in dialogue-based interaction [6]

3 METHODOLOGY

We will now describe the salient details of our hypotheses, experimental design, procedure, and measurements. We refer the reader interested in an in-depth of our hypotheses and work on Indirect Speech Acts to our previous work [1].

3.1 Hypotheses

Previous research has suggested that robots' size, imprecision of movement, and *actual* impoverished capabilities all contribute towards *perceived* impoverished capabilities [2, 4, 13]. We expect this to be especially true for suit-controlled teleoperated humanoid robots like the one used in our experiment, which are disadvantaged on all of these fronts relative to their human teammate. We thus hypothesized **(H1)** that both autonomous *and teleoperated* robots would be perceived as less intelligent and capable than humans performing equivalent tasks.

And because previous research has suggested that higher levels of robot autonomy correlate with higher levels of blame and scrutiny [2, 5], we hypothesized **(H2)** that teleoperated robots would be perceived as more successful after completing a task than would an autonomous robot.

3.2 Experimental Design

We conducted a laboratory study in which participants each taught a new skill to a **human** learner and to a **robot** learner. Participants were either told that the robot learner was **teleoperated (TC)** or **autonomous (AC)**. We used identical Wizard-of-Oz interfaces in both conditions, and in neither condition did any participant see nor know anything about the VR teleoperation interface or the teleoperator. Thus, any differences in perception of the robot existed solely in participants' minds. We would expect participants in the **TC** condition to assume the teleoperator to have identical cognitive capabilities and reduced physical capabilities to a co-present human, and participants in the **AC** condition to assume the robot to have reduced cognitive *and* physical capabilities to a co-present human.

Participants interacted with each agent by teaching them how to complete a task in a room divided into two areas: a *teaching* area in which the participant was seated, and a large *experiment* area. Participants were seated in front of a *diorama* which replicated the experiment area in miniature (Fig. 1). Both the diorama and experiment area were divided into four quadrants containing a variety of objects: four cardboard boxes (each labeled with a different letter), and three colored towers (Lego blocks in the diorama and aluminum cans in the experiment area).

Participants were told to arrange their diorama *however they wished* (except that cans could not be translated, and boxes could not be flipped over or stacked), after which they would be tasked with teaching a human or robot agent how to replicate that arrangement using the full-size objects found in the experiment area. In **AC**,

participants were told that if the learner was a robot, it would be autonomous; in **TC**, participants were told that if the learner was a robot, it would be a teleoperated robot controlled by a human who, using an interface, could make the robot say a limited number of things. After participants finished arranging their diorama however

Figure 1: Experiment room, including diorama.

they wished, the researcher retrieved the first agent. The agent moved in front of the participant, introduced themselves, and stated "Today I will be listening to your instructions to arrange this room in the manner you have done here," gesturing towards the participant's diorama. Participants then gave instructions to the agent to arrange the full-size objects to replicate their diorama arrangement, and the learner complied to the best of their ability. Both human and robot learners restricted their utterances to "Okay", "Yes", and "No" whenever possible. The robot's utterances were selected by a human confederate, and synthesized using a text-to-speech interface. Once the interaction was over, the agent said "Goodbye" and exited.

3.3 Teleoperation Interface

There were several key decisions made as part of this work's experimental design in order to facilitate ease of interaction, and minimize variability between participants. First, in order to ensure participant engagement, naturalness of interaction, and prevent bias of the experimenter on participants' utterances, participants in this experiment had *free control* over their arrangements and how those arrangements were described (c.f. [9, 10]). Second, a major key to this experiment was the specific robot and the control interface

Figure 2: Teleoperation exo-suit and humanoid robot.

used: a humanoid robot and a one-to-one exo-suit developed by Kindred, as seen in Fig. 2.

This robot, made with a combination of 3D printed joints and limbs connected by hobbyist servos, has 5 degree-of-freedom (DoF) arms, 10 under-actuated fingers (a plastic prosthetic hand model), and a 2 DoF head. The robot's torso contains a small computer and WiFi adapter for wireless communication, and its' base contains a battery to power the robot for several hours untethered.

Importantly, the robot and exo-suit were coupled with an Oculus Rift virtual reality headset so that a human operator could be visually immersed in the robot's environment, using the stereoscopic cameras on the robot as the video feed for the human operator. Additionally, two microphones on the robot's head delivered stereophonic auditory data so that language and sounds could be spatially localized even if the audio source was not in the robot's line of sight. A set of 3 pedals were used to move the 4-wheeled base of the robot.

Lastly, to standardize language and voice used throughout the study, the robot's voice output was controlled by a second operator using a limited text-to-speech interface which allowed the robot to utter the same introductory phrases and limited responses used by the human confederate.

3.4 Experimental Procedure and Participation

Participants first completed a questionnaire gathering information on participants' demographics and previous experience with robots. All questionnaires were carried out in a separate survey room. Participants then moved to the experiment room and conducted the main task. Next, participants completed a questionnaire assessing their perceptions of the agent and the success of the task. After completing this questionnaire, participants were told they would perform the same task with "another agent" and that they were again free to arrange their diorama however they wished. Participants were not told what type of learner they would interact with next, but in all cases the type of learner varied to counter the first interaction; if a robot learner was used in the first experiment, a human learner was used in the second, and vice versa. Upon finishing the second interaction, participants answered the experiment questionnaire again, as well as additional questions comparing the two tasks and agents.

Thirty-three students and university employees were recruited through fliers and university class forums. Participants (21 Female, 12 Male) ranged in age from 18 to 25 (M=20.85, SD=1.37). All participants were given $10 as compensation for their time.

3.5 Measurement

In addition to the demographic survey, as previously described, questionnaires were used to assess participants' perceptions of their human and robot teammates. Participants' perceptions varied along a variety of dimensions, including cooperativity, capability, annoyingness, creepiness, and responsiveness

In addition to these subjective, self-reported measurements, we also collected several objective behavioral measures. For each participant, video recording was used to assess the number of words used by participants, and the percentage of utterances used by participants that were ISAs, as well as the complexity of participants' arrangements, as described in our previous work [1].

4 RESULTS

In this section we provide a brief summary of our experimental results. We refer the interested reader to our full results described in our previous work [1].

Participants used significantly more words to describe the task in the **TC** (M=204.88, SD=181.60) than in **AC** (M=72.82, SD=42.83), $F(1,29)=8.05$, p=.008. An ANCOVA using task complexity as a covariate attenuated this effect, $F(1,29)=4.43$, p=.02. Additionally, participants used more words to describe the task (to either learner) in **TC** (M=153.28, SD=103.40) than in **AC** (M=80.09, SD=40.02), $F(1,29)=6.91$, p=.014, an effect attenuated in a subsequent ANCOVA, $F(2,62)=4.72$, p=.012. An interaction between condition and learner was also found: far more words were used with robots in **TC** (M=204.88, SD=181.60) than with *humans* in **TC** (M=101.69, SD=44.79), robots in **AC** (M=72.82, SD=42.83) or humans (M=87.35, SD=46.27) in **AC**, $F(1,29)=7.79$, p=.009, an effect that was strengthened in a subsequent ANCOVA, $F(6,58)=3.25$, p=.008.

Furthermore, a number of significant differences were found between perceptions of the human learners and the robot learners:

	H(M, SD)	R(M, SD)	F(1,29)	p
Capable	6.88, 0.33	5.97, 1.04	28.07	<.001
Annoying	1.36, 0.93	2.00, 1.17	7.02	.013
Creepy	1.67, 1.08	2.73, 1.72	11.16	.002
Conscious	6.97, 0.17	3.97, 1.78	91.63	<.001
Easy Interaction	2.06, 1.84	3.15, 1.73	6.67	.015
Comprehension	6.76, 0.56	5.94, 0.66	29.45	.007
Understanding	6.79, 0.48	6.27, 0.76	12.96	.001
Gaze Following	5.42, 1.85	3.70, 1.84	17.07	<.001
Perceived Success	6.78, 0.48	6.12, 0.93	12.99	.001

5 DISCUSSION

Throughout conducting this experiment and by way of using the exo-suit teleoperation interface for numerous hours, a number of challenges and learnings were made. Firstly, given that our experiment required real time remote communication via motor control, vision, and hearing, we found any spikes in latency to be extremely detrimental to the running of the experiment. A spike in latency might lead to: a participant's instruction not being heard, inaccurately executing a movement, or having to operate blindly (depending on the modality affected). With the technology used in this study, latency was not generally an issue unless there was high bandwidth usage or congestion on back-end server calls. Future work might mitigate this issue by ensuring that high quality WiFi adapters are used, and that high bandwidth internet proliferates the experiment area.

Secondly, the size of the items we chose to move around the room were slightly too large in proportion to the size of the robot, which can be gleaned from Fig. 2. Although this choice was deliberate (wanting to maximize the size of items used by the robot so they wouldn't be trivial for the human to also use), it occasionally resulted in the teleoperater being unable to view the surroundings while controlling the robot, as the robot's peripheral vision might have been obscured by the item.

Finally, there were a number of lessons learned about how this teleoperation mechanism affected our experiment design. Because

we were using the same teleoperator, exo-suit, and robot for all of our trials, and because we had the capability of seeming exceptionally human-like (with fluid, complex movements, or advanced language), we had to ensure that in any and all trials the robot did not seem too human-like to be obviously perceived as teleoperated, nor too robotic to be obviously perceived as autonomous. Thus, the teleoperator used standardizing strategies such as keeping their arms on arm rests whenever possible, so that complex forearm movement did not expose the level of autonomy. Additionally, the text-to-speech interface and set of commands was used to mitigate any variability in language across participants, but this prescribed language did seem to force the human-human interactions to be more awkward than they would have been with more fluid language.

These challenges aside, we believe that this research could not have been possible without a teleoperation mechanism similar to the one used, as any limitations to dexterity or immersion would have left the teleoperator detached from the interaction, and thus unable to respond to unpredictable participant requests in real-time, and unable to succeed at task completion.

We hypothesized **(H1)** that not only would robots described to participants as *autonomous* be perceived as less intelligent and capable than humans would be, but also that robots described to participants as *teleoperated* would be perceived with similarly diminished capabilities, even though in reality all three learners had identical cognitive capabilities. In fact, this is just what we observed. Our results showed that participants rated both autonomous *and teleoperated* robots as less understanding of their instructions and less likely to understand high level commands. This is striking, as such ratings should depend only on the *mental faculties* of the learner, and yet, participants rated the human-teleoperated robot no differently in this respect than they did in **AC**. Furthermore, both autonomous *and teleoperated* robots were rated as more annoying, creepy, harder to interact with, and overall less capable *and conscious* than their human counterparts.

This suggests that regardless of whether or not a robot is human- or AI-controlled, humans are likely to see the robot's form as hindering its controller's capabilities *and intelligence*. This is particularly significant for human-robot collaboration, as it suggests that people may view not only a teleoperated robot, but also *its teleoperator*, as inferior to a present human counterpart, altering both social dynamics and expectations of success.

Finally, because higher levels of autonomy have previously been correlated with higher levels of blame and scrutiny, we hypothesized **(H2)** that autonomous robots would receive less credit for successful completion of the task than would teleoperated robots (i.e., that teleoperated robots would be rated as more successful). While we did not find evidence supporting this hypothesis, we did observe that participants in **TC** retrospectively judged the arrangements provided to robot learners to be more complex than did participants in **AC**, even when controlling for the actual complexity of their arrangements. This suggests that participants may have attributed more of the credit for task success to the *learner* in **TC**, but to have retrospectively assumed *simplicity of arrangement* in **AC**, which would indirectly support H2.

6 CONCLUSIONS

In this paper, we investigated the differences in how humans instruct both humans and robots when choosing their own task,

particularly examining the differences between instructions given to purportedly autonomous and teleoperated humanoid robots controlled through identical immersive virtual reality interfaces in which teleoperator motions are replicated by the robot in real time.

Our results suggest variations in how human teammates, autonomous robot teammates, and teleoperated robot teammates were perceived. Specifically, our results suggest that human-teleoperated robots were perceived as less intelligent than human teammates; a finding with serious implications for human-robot team dynamics.

Additionally, we describe how this research was made possible using a VR suit-based teleoperation interface, and the lessons learned with the design and use of that interface.

Future research should investigate (1) how the use of a VR suit-based teleoperation interface directly compares to use of a GUI-based or joystick-based teleoperation interface in controlling a humanoid robot; (2) differences in gaze and gesture patterns accompanying the instructions given to humans and both autonomous and teleoperated robots; (3) what aspects of a teleoperated robot's appearance and behavior contribute to the decreased perceptions of its teleoperator's intelligence and consciousness.

ACKNOWLEDGMENTS

REFERENCES

[1] Maxwell Bennett, Tom Williams, Daria Thames, and Matthias Scheutz. 2017. Differences in Interaction Patterns and Perception for Teleoperated and Autonomous Humanoid Robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.
[2] Elizabeth Cha, Anca D Dragan, and Siddhartha S Srinivasa. 2015. Perceived robot capability. In *24th Int. Sym. on Robot and Human Interactive Comm.*
[3] Paulo G. de Barros and Robert W. Lindeman. 2008. A Survey of User Interfaces. (2008).
[4] Michael A Goodrich, Jacob W Crandall, and Emilia Barakova. 2013. Teleoperation and beyond for assistive humanoid robots. *Reviews of Human Factors and Ergonomics* 9, 1 (2013), 175–226.
[5] Taemie Kim and Pamela Hinds. 2006. Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. In *15th Int. Sym. on Robot and Human Interactive Communication*.
[6] Bilge Mutlu, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *4th Int'l Conf. on Human-Robot Interaction*.
[7] Erhan Oztop, David W Franklin, Thierry Chaminade, and Gordon Cheng. 2005. Human–humanoid interaction: is a humanoid robot perceived as a human? *International Journal of Humanoid Robotics* 2, 4 (2005).
[8] I. Rodriguez, A. Astigarraga, E. Jauregi, T. Ruiz, and E. Lazkano. 2014. Humanizing NAO robot teleoperation using ROS. In *Int'l Conf. on Humanoid Robots*.
[9] Stephanie Schreitter and Brigitte Krenn. 2014. Exploring inter-and intra-speaker variability in multi-modal task descriptions. In *23rd Int. Sym. on Robot and Human Interactive Communication*.
[10] Stephanie Schreitter and Brigitte Krenn. 2016. The OFAI Multi-Modal Task Description Corpus. In *10th Int'l Conf. on Language Resources and Evaluation*.
[11] Christopher Stanton, Anton Bogdanovych, and Edward Ratanasena. 2012. Teleoperation of a humanoid robot using full-body motion capture, example movements, and machine learning. In *Aus. Conf. on Rob. and Automation*.
[12] Mike Stilman, Koichi Nishiwaki, and Satoshi Kagami. 2008. Humanoid teleoperation for whole body manipulation. In *Int. Conf. on Rob. and Autom.*
[13] Astrid Weiss, Daniela Wurhofer, Michael Lankes, and Manfred Tscheligi. 2009. Autonomous vs. tele-operated: How people perceive human-robot collaboration with HRP-2. In *4th Int'l Conf. on Human-Robot Interaction*.
[14] Zhengyou Zhang. 2012. Microsoft kinect sensor and its effect. *IEEE Multimedia* (2012).